Outsourcing Concepts: Deference, the Extended Mind, and the Expansion of our Epistemic Capacity

Cathal O'Madagain

Forthcoming in *Socially Extended Knowledge*, Andy Clark, J. Adam Carter, Jesper Kallestrup, S. Orestis Palermos, Duncan Pritchard (eds). Oxford University Press

Abstract: Semantic deference is the apparent phenomenon whereby some of our concepts have their content fixed by the minds of others. The phenomenon is puzzling both in terms of *how* such concepts are supposed to work, but also in terms of *why* we should have concepts whose content is fixed by others. Here I argue that if we rethink semantic deference in terms of extended mind reasoning we find answers to both of these questions: the minds of others can be understood to play a role in storing the semantic knowledge underpinning our concepts without undermining their functionality, and this 'outsourcing' of semantic knowledge greatly expands our overall knowledge-bearing capacity, both at the level of the individual and the community. I conclude that deference in fact affords stronger grounds for a social expansion of knowledge than standard extended mind reasoning permits.

1. Externalism, Passive and Active

Suppose Ned asserts the following:

1) Narcolepsy is a disease of the nose

If Ned asserts 1), then we'll likely attribute to Ned a false belief about narcolepsy, that it is a disease of the nose. But this leads to a puzzle. Standard theories of concepts assume that the content of a concept is fixed by internal states of the bearer – an internally stored theory (Gopnik and Welman 1992), prototype (Rosch and Mervis, 1975), or exemplar (Medin and Schaffer 1978) of the concept. Ned's internal states, however, appear to identify something *other than* narcolepsy as the reference of his NARCOLEPSY concept. His internal states, after all, tell him that narcolepsy is a disease of the nose. Cases like this lead Putnam (1975) and Burge (1979) to propose that the content of at least some of our concepts must be fixed by facts about *others* – a view known as 'social externalism' about concepts. The reference of Ned's NARCOLEPSY concept is, on this view, fixed by the beliefs of his doctor, or the medical experts in his community, to whom Ned 'defers' on the content of that concept.

Now suppose that a reliable Leipziger tells me:

2) The Linden trees in Leipzig are lovely

Upon hearing 2), it would seem that I can come to know that the Linden trees in Leipzig are lovely. Knowledge acquired by testimony appears routine among our ordinary knowledge attributions, and denying that we can acquire knowledge in this way would drastically curtail the knowledge we generally take ourselves to have (Lackey 2007, Pritchard 2010). But it would also seem that such acquisition of knowledge is possible even if one is not certain of the meaning of all the words in that testimony. I might after all hear 2) before I get to Leipzig, and having arrived, ask around to see which ones are the Linden trees so that I see their loveliness for myself. This seems to entail both I did not know exactly which the Linden trees were, but did know that they were lovely – in other words, that I acquired a piece of knowledge by testimony involving a concept I didn't fully grasp. Goldberg (2009) argues that cases like this also motivate semantic deference: the term 'Linden trees' in the belief I acquire has its reference fixed by the belief of the Leipziger who presented me with this testimony.

We have two reasons, then, to think that semantic deference may apply to many of our concepts. First, it seems that ordinary practices of belief attribution lead us to attribute beliefs to others who clearly don't grasp some of the concepts in those beliefs. Second, it seems that we often learn things by testimony even when we don't fully grasp all the concepts in that testimony. Together these suggest that some of our beliefs are governed by a principle we might call 'deference':

Deference: A belief is deferential if the content of some of the terms in that belief is fixed by the minds of others

For many, however, the idea that *Deference* might govern some of our beliefs raises more questions than it answers: specifically a 'how' question, and a 'why' question.

The first is *how* could Ned really be thinking of narcolepsy if he cannot identify it himself. A long-standing tenet in the philosophy of mind, after all, is that it is a precondition on having a thought about something that one can identify that thing, something Evans called 'Russell's principle' (Russell 1921: 58, Evans 1982: 89). Alternatively put, to be attributed a concept one must have what we might call the 'semantic knowledge' that fixes the reference of the concept – one must know what the concept refers to either by description, or be able to recognize that thing. And yet for deferential concepts, it would seem that the thinker does not have the semantic knowledge behind the concept, and cannot therefore identify what those concepts refer to. The standard way to makes sense of this, sometimes called the 'metalinguistic' account, is to suppose that Ned's NARCOLEPSY concept is equivalent to something like 'whatever the experts around here mean by 'narcolepsy'' (Searle 1983, Lewis 1997, Jackson 1998, Chalmers 2006). As we shall see, however, this account is problematic.

A further question is *why* would we have deferential concepts. Even if we can somehow make sense of how an agent could have a belief with deferential content, those beliefs will clearly not be as efficient as beliefs without deferential content. To find out what 'narcolepsy' means Ned will have to talk to his doctor, if his belief is deferential. And that's much more trouble than Ned would have to go to if he could tell what the terms in his belief picked out unaided. But concepts

and beliefs take up cognitive space – we have to store them in memory. So surely Ned would be better off not trading in clunky deferential concepts, but only in non-deferential concepts, and using his cognitive space for concepts and beliefs that he doesn't need others to understand. From a general utility point of view, deference seems mysterious. And if deference is pervasive – so that many of our concepts are deferential – then so much more the mystery. The standard account has little to say about this second question.

Both the 'how' and 'why' questions can be understood to stem from our basic functional understanding of beliefs, of which concepts are the constituents. The functional definition of a belief, after all, is generally thought to be that of translating desires into action (Putnam 1963, Prinz and Clark 2004). For example, if I believe ARTICHOKES CURE ARTHRITIS, and I desire to cure my dose of arthritis, then my belief tells me what to do – go get some artichokes. This functional understanding of a belief is captured in the following principle:

Utility: A cognitive state is a belief if it can be used to translate our desires into action

The reason semantic deference is so puzzling, I suggest, is that it seems like there is a straightforward conflict between *Deference* and *Utility*. If *Deference* governs some of our beliefs, then some of our beliefs may contain concepts we don't fully grasp. If I don't fully grasp the concepts in my beliefs then it would seem that they could not play this basic functional role. Having the belief ARTICHOKES CURE ARTHRITIS seems to be of little use to me to address my desire to cure my arthritis, if I don't know what ARTICHOKES means. If a cognitive state is deferential, it therefore seems like it could not satisfy *Utility*, and as a result hardly count as a belief in the first place.

Here I argue that answers to both the 'how' and 'why' questions, and thereby a resolution to the conflict between *Deference* and *Utility*, becomes available if we explore semantic deference in terms of the reasoning behind what is generally thought of as another kind of externalism entirely. This is the so-called 'active externalism' of Clark and Chalmers' (1998) 'extended mind' theory. Social externalism is generally thought of as a kind of 'passive externalism', according to which aspects of our environment are not actually doing any cognitive work, but simply altering the reference of our thoughts. According to the thesis of the extended mind, on the other hand, aspects of our environment *actively support* our cognitive processes. Suppose Inga and Otto both take a walk to the museum, and Inga remembers the route using her biological memory, but Otto takes down the route in a notebook. The next day, both independently make their way to the library – Inga using her biological memory, Otto using his notebook. If we individuate cognitive types functionally and allow for multiple realizability, so that we do not restrict cognitive vehicles to particular kinds of material, we should regard Otto's notebook as a part of his memory, and hence his mind: it performs, after all, exactly the same function as Inga's biological memory. The only major difference is that the information accessed by Otto is stored outside of his head, but the information accessed by Inga is stored inside. Unless we are to beg the question against the possibility that our minds might extend into our environment, we cannot use this difference to reject that the cognitive types are the same. Given this 'functional parity' reasoning, all sorts of

ways in which our minds extend into and are augmented by our environment become apparent. But the view also carries with it a striking rationale – if we can use aspects of our environment in our cognitive processes, then we can harness our environment to dramatically augment our cognitive capacities, and this explains *why* we might operate with cognitive processes that extend into our environments.

In the next section I turn to the standard 'metalinguistic' account of deference, and consider the problems of this approach. In section 3, I explore how extended mind reasoning can be used to provide an alternative answer to the 'how' question that faces none of the problems of the metalinguistic approach, and in section 4, I explore how it can also provide us with an answer to the question why our thoughts might contain deferential concepts.

2. The Metalinguistic Account

On the metalinguistic account, we make sense of *how* Ned could be thinking about narcolepsy by supposing that the term 'narcolepsy' in his thought really amounts to something like 'whatever experts means by 'narcolepsy''. This certainly seems to hit the right referential target. Suppose Ned's concept is deferential to the local expert's, then his NARCOLEPSY concept will have the same content as the expert's NARCOLEPSY concept, and the metalinguistic description surely identifies what the expert's NARCOLEPSY concept refers to. But there are problems with the story.

The first is that it commits us that beliefs with deferential content are in part about words. Ned must be thinking about the *word* 'narcolepsy', when he is entertaining or expressing his belief 'narcolepsy is a disease of the nose', if the term 'narcolepsy' in that thought is equivalent to a metalinguistic description. This is doubtful for several reasons (versions of which are discussed early on by Burge 1979: 93ff). First, if we were to ask Ned whether he was thinking about the word, it is surely conceivable that he would deny it:

> The subject may maintain that his reasoning did not fix upon words. He may be brought up short by a metalinguistic formulation of his just-completed ruminations, and may insist that he was not interested in labels (Burge 1979: 97)

So the story conflicts with the content of the beliefs we would attribute to ourselves. Since ordinary practices of belief-attribution are what motivate semantic deference in the first place, that seems like a big problem.

Second, it seems possible that someone could acquire a deferential belief like 1) or 2) without having concepts of MEANING or REFERENCE at all. Young children, for example, might acquire many beliefs by testimony without fully grasping the concepts in those beliefs – coming to believe that electricity is dangerous, or that the stove runs on gas, without having the ability to uniquely identify either electricity or gas. But it seems highly doubtful that they should grasp the concepts MEANING or REFERENCE for this to take place.

Finally, this same feature of importing words into Ned's thought has an extremely odd implication for the relationship between Ned's thoughts and the expert to whom he defers. Ned and the expert can, after all, surely entertain the

same thought content about narcolepsy – Ned taking it as true, the expert taking it as false. But if this means that Ned is having a metalinguistic thought, then the expert must also be having a metalinguistic thought, if they are really entertaining the same thought. And this means the expert is thinking about narcolepsy as 'whatever I mean by 'narcolepsy'', or worse, 'whatever the expert means by 'narcolepsy'', where 'expert' refers to herself. Either seems absurd.

A second major problem for the metalinguistic account is it seems to generate the attribution of beliefs that clearly are not in a position to satisfy *Utility*. The basic principle behind the metalinguistic approach appears to be that a belief of mine can acquire content C if someone else's utterance has the content C, and I entertain a thought about the content of that utterance we might say 'quotationally' – that is, believing 'whatever she said'. Sometimes this produces convincing results. For example, suppose that I have arthritis and I want to get rid of it, and a neighbour tells me:

3) Arthritis can be cured with artichokes!

Now suppose that I don't know exactly what artichokes are, beyond knowing they are some kind of vegetable, and I defer to the speaker on the precise individuation of artichokes. On the metalinguistic view, the belief I acquire amounts to something like ARTHRITIS CAN BE CURED BY WHATEVER MY NEIGHBOUR MEANS BY 'ARTICHOKES'. Is this belief of any use to me? It seems clear that it is. I can use this belief to address my desire to cure my arthritis, simply by recognizing that it bears on this desire, and that artichokes are some kind of vegetable. I can then head to the grocer and ask 'which ones are the artichokes please'. The grocer will then point me to the artichokes, whereupon I can buy some and address my desire to cure my arthritis. So, it would seem that some cases of belief acquisition involving metalinguistic content can indeed satisfy *Utility*.

But now suppose that I meet another neighbour, this time an Uzbek who I consider reliable, and she says to me:

4) Artrit artishok bilan davolash mumkin

As it happens, what the Uzbek has said is that arthritis can be cured with artichokes. If I take her to be reliable, I can certainly form the belief that what she said is true. It is nevertheless obvious that the belief that I acquire here, that 'whatever the Uzbek meant by 'Artrit artishok bilan davolash mumkin' is true', does not amount to my acquiring a belief with the content that artichokes cure arthritis. If I believe that I have arthritis and desire to find a cure, believing that 4) is true is not going to be of any help to me in addressing that desire. But the 'quotational principle' seems to allow that this is indeed the content of the belief I have acquired, or at least doesn't tell us why not. So the metalinguistic approach appears to attribute to us beliefs that clearly fail *Utility*.

Apart from these problems for a metalinguistic account of *how* deferential beliefs are supposed to work, note that the metalinguistic account offers us really no explanation at all for *why* we would possess deferential concepts. Even if we can show that in some cases a metalinguistic description will offer us a functional belief, these beliefs are clearly less efficient than non-deferential beliefs, since in order for me to fully translate these beliefs into action I will have

to rely on the availability of someone else – an expert or the like. And the metalinguistic account gives no hint as to what advantages might exist in deference to offset this disutility.

Extended mind reasoning, however, offers us an alternative approach.

3. Extended Cognition and the How Question: Outsourcing Concepts

The basic move at work in extended mind reasoning is to identify cases of functional parity between cognitive states or activities that rely on features of the thinker's environment, and those that don't. Once we establish functional parity, we are in a position to show, assuming a functionalist approach to cognitive states in general, that some of our cognitive states extend into our environment. In this section I want to show that the case for functional parity can be made for beliefs that include deferential concepts, and that this offers us an answer to the 'how' question that does not run into the problems faced by the metalinguistic account.

There are two ways in which we might be concerned that deferential and non-deferential thoughts differ, sufficient for us to doubt their functional parity. The first is in a non-deferential case, the concept bearer has access to the semantic knowledge that fixes the reference of her concepts either by introspection or recognition. In a deferential case, the putative concept bearer has to interact with someone else to access that semantic knowledge. The second is that this due to this difference, the retrieval of semantic knowledge behind a deferential concept could be indefinitely delayed. I'll look at each in turn.

First let's consider what is involved in retrieving semantic knowledge in a non-deferential cases. Perhaps, we might suppose, that knowledge is transparently available to us. Whenever I have a thought about horses, we might imagine, I can tell what I am thinking about by introspecting, and along will trot a theory, prototype, or exemplar of horses to tell me what my thought is about. But although such introspection of conceptual content is surely possible, and perhaps even necessary in at least some cases (Jackson 1998, Chalmers 2006), it is equally certain that it doesn't always work.

That it doesn't always work is illustrated by various well-studied phenomena of 'meta-memory' (Flavell 1976). One is the 'tip of the tongue' phenomenon – where we are sure that we have a piece of information stored away in our minds, but it takes some effort to retrieve it. This is something we often encounter in games of trivia. The information subject to tip-of-the-tongue phenomenon is often a word ("What's the former name of Tanzania? I'm sure I know this… it's…. Tan… Tanganyika!") (Brown and McNeill 1966, Schwarz and Metcalf 2010); but analogous phenomena apply to propositional knowledge, where we encounter 'feelings of knowing' something without immediately finding the knowledge ("who shot Tony Soprano? I'm sure I know this… it was… Uncle Junior!")(Koriat 1995, 2000); and of course this can also apply to semantic knowledge ("what is an echidna… I'm sure I know… it's… a spiny hedgehog!"). All this provides evidence for what we might call the temporary opacity of semantic knowledge: it is possible to have a piece of semantic knowledge, and know that one has that piece of knowledge, without having immediate access to the knowledge – the semantic knowledge and the knowledge that one has that knowledge can be 'decoupled'.

What must the relationship between meta-memory and memory be, that such decoupling is possible? It would seem that some sort of fallible mechanism must interface between the occurrence to us that we have a piece of information in our memory, and the actual retrieval of that information from memory. Since some sort of fallible retrieval mechanism is necessary in order for us to retrieve information from biological memory, then the fact that information stored in an external source requires us to employ a fallible mechanism does not raise as much of a problem for regarding the two to be functionally similar as we might have expected. If we need to deploy a lookup procedure that involves consulting a directory in handheld device (Clark 2003: 42; Clark 2008: 80), or using a shared language with a trusted partner (Wegner 1986, Tollefsen 2006), these procedures can be seen as functionally similar to the kind of procedure required to retrieve information from biological memory. The difference in means of retrieval of semantic knowledge behind deferential and non-deferential concepts should not, therefore, undermine their functional parity.

However, the mere retrievability of this semantic knowledge is not enough to show that it is playing a functional role in our mental states. If that were the case, then since I can retrieve the entire contents of the local library with a little footwork, then it would seem that extended mind reasoning would imply that I already know everything that I might learn with a trip to the library. This seems absurd (Clark 2008: 80; Varga 2013; Marsh, 2010; Rupert, 2009; Allen-Hermanson 2012). To be on functional par with knowledge in my memory, a piece of externally stored knowledge must in addition to being retrievable, be integrated with my cognition in the same way in which locally stored knowledge is. And here we might worry that the second difference between deferential and non-deferential thoughts will become a problem – that the retrieval of semantic knowledge in deferential cases could be greatly delayed.

Clark and Chalmers' 'Glue and Trust' conditions are designed to say what it means for a piece of information to be sufficiently integrated in a cognitive system to play the functional role of a belief. The conditions are that the information should be i) 'typically invoked', ii) 'easily accessible', iii) 'automatically endorsed' and iv) 'consciously endorsed at some point in the past' (Clark 2008: 79). We do, after all, typically invoke the information stored in our biological memory, have easy access to it, automatically endorse it, and have endorsed at some point in the past – when we acquired it originally. And it is clear that these features bear on their functional efficacy. If we did not automatically endorse our own beliefs, but having retrieved one from memory found ourselves wanting to verify if once again, then it would not be functionally effective – since we would have to reassure ourselves of everything we had once gone to the trouble of learning whenever it was recalled. That would seem to defeat the purpose of learning in the first place. And these conditions do indeed seem to block cognitive bloat. They rule out the contents of the local library, after all: I can retrieve this information, but I do not automatically endorse it, have not endorsed it in the past, and it is certainly not as easily accessible as my locally stored memories. However, in constraining the attribution of functional roles to information by the Glue and Trust conditions, we may have also ruled out the semantic knowledge stored in the minds of others.

So, do the Glue and Trust conditions apply to semantic knowledge? The first and third of these should not raise too much of a concern. Deference more

or less entails that the holder of a deferential concept always and automatically accepts the authority of the person to whom they defer. Now let's consider the fourth constraint. Must I have consciously endorsed in the past the semantic knowledge underpinning my concepts in order for that knowledge to fix the reference of those concepts? Although this may be a plausible constraint on something counting as a piece of propositional knowledge of mine, it does not seem to apply to semantic knowledge. Simply put, it hardly makes any sense to say that I have consciously endorsed that my concept HORSE refers to horses – or that a child who has acquired the concept RED has at some point consciously endorsed that her concept RED picks out red things. She likely has acquired the concept without having any thoughts about concepts – as discussed, she might not even be in possession of the concept CONCEPT. So while having beliefs might be something that involves conscious endorsement at some point, having concepts does not.

The second constraint, however, addresses exactly the worry for deferential content raised above. This is that we should have ready access to content stored offline for it to count as contributing to the functionality of our beliefs. Obviously we do not have immediate access to the semantic knowledge of grocers when we entertain thoughts about artichokes. If the immediate availability of semantic knowledge is necessary for it to play a functional role in a belief in ordinary cases, then deferential cases might not be on functional par with non-deferential cases.

In fact, however, we have good reasons to think that such immediate access could not be a requirement on entertaining a functionally effective belief in ordinary cases. First, there are the considerations that derive from the arguments from ignorance and error behind classic natural kind externalism (Putnam 1975, Kripke 1980). It often appears to be the case that we cannot produce via introspection, even with all the time in the world, a uniquely identifying description of the reference of our concepts, and those that we are inclined to produce might often be incorrect. This consideration has convinced many that by and large our ability to identify the reference of our concepts will much of the time be recognitional at best – that for the semantic knowledge to be retrieved, I will need to have perceptual access to the thing I'm thinking of, whereupon I can identify it (Evans 1982, Sterelny 1983, Peacocke 1992). But of course it is not the case that we have to be immediately engaged in recognition of the reference of the concepts for them to play a role in our thoughts – I don't need a horse to walk into the room to think about horses. The upshot is that, if the semantic knowledge behind even some of our concepts is recognitional, then it could not be the case that we need immediate access to the semantic knowledge behind the terms in our beliefs for the belief to be functionally effective.

On the other hand, we might be strongly committed to the introspective availability of the semantic knowledge behind of our concepts, as the neo-Fregean program seems to be. But here too it is doubtful that this commitment requires that this semantic knowledge must be *immediately* available. One reason to doubt this comes from considerations of efficiency. Psychological theories of concepts that are consistent with a Fregean approach, such as prototypes, exemplars, or theories, generally take these internal states to be extremely complicated. If it were necessary to retrieve the full content of a

prototype behind my CAT concept every time I were to think THE CAT IS ON THE MAT, this would be extraordinarily inefficient (Wicklegren 1979, 1992, Mendelovici 2010). Furthermore, one might take the phenomenal character of a thought to be a significant indicator of its role in our cognition, a view that is center stage in the 'phenomenal intentionality' research program (Kriegel 2013) and which is often partnered with a neo-Fregean account of content (e.g. Searle 1991, 1992). But since it is not a part of the phenomenology of thoughts that our concepts occur to us in fully 'unpacked' form then this too speaks against attributing to that unpacking any essential immediate role in our thoughts.

It is not at all obvious, as a result, that we need to have ready access to the semantic knowledge underlying a concept in order for a belief involving that concept to be functionally effective – to satisfy *Utility*. If I believe that artichokes cure arthritis, and I desire to cure arthritis, then this belief must lead me *ultimately* to finding some artichokes to use to cure my arthritis if the utility of the belief is to be fully realized. But the semantic knowledge that allows me to identify the artichokes – perhaps the capacity to recognize artichokes when I see them – needn't be recruited at the first stage of this process, where I simply recognize that the belief ARTICHOKES CAN CURE ARTHRITIS is relevant to my desire I WANT TO CURE MY ARTHRITIS. If I can in the short term recall that artichokes are 'some vegetable', then the occurrence of this thought will be sufficient to get me to head to the grocery store where I can fully retrieve the semantic knowledge underpinning the term, which may involve actually seeing the artichokes so that I can recognize them. Since a delay in the retrieval of semantic knowledge for a non-deferential belief does not interfere with its good functioning, a delay brought about in the case of a deferential belief due to the temporary unavailability of an expert does not mean the deferential belief is not on functional par with the non-deferential one.

Understood in this way, the semantic knowledge in the mind of the expert us playing a role in the ultimate functionality of my belief. But this means that a part of the cognitive process that supports the functionality of my belief takes place outside my mind – the retention of the semantic knowledge, which if it took place in my brain, we would not hesitate to regard as a cognitive process. And so we are left with semantic deference assessed as a straight-forward instance of extended cognition: in the non-deferential 'artichokes' case, the internally stored cognitive ability to identify artichokes plays a crucial role in the realization of the function of that belief; in the deferential case, that cognitive ability is the expert's.

This gives us an extended cognition answer to the 'how' question, and thereby a partial resolution of the conflict between *Deference* and *Utility*. And it is a resolution that avoids the problems of the metalinguistic account. First, once we take the functionality of a deferential thought to be realized in part by the retrieval of knowledge from another mind, we no longer need to suppose that thoughts about words occur in deferential thoughts. Non-deferential thoughts, after all, involve no reference to the sub-personal mechanisms that make the retrieval of semantic knowledge possible. Similarly, although it must be possible for us to know how to get that knowledge in a deferential case – let's say by asking the person we defer to – there is no longer any pressure to suppose that thoughts about that mechanism are involved in the original occurrence of the deferential thought. Second, since we have not had to appeal to the 'quotational'

principle that the metalinguistic account must rely on, we do not find ourselves committed to the attribution of dysfunctional beliefs like 4).

A further question, however, remains: even if beliefs with deferential content can achieve functional parity with beliefs with non-deferential content, it is clear that they are *far less reliable*. That is, more things are likely to go wrong for a belief when the semantic knowledge behind some of the concepts in that belief are stored in other people's minds – I might not be able to find someone with the knowledge, their memory might fail even though mine does not, etc. Although deferential beliefs may be able to play the functional role of non-deferential beliefs, they appear to be doing a worse job, and so we might wonder what the point is of having them. Extended mind reasoning, however, offers us an answer to this question too.

4. Extended Cognition and the Why Question: Expanding our Epistemic Capacity

To address the 'why' question, let's first turn to the literature on the distribution of cognitive labor, a close ally of extended mind reasoning. One of the basic insights here is that if many people divide the work of a cognitive task, the overall efficiency in performing that task is greatly increased (Kitcher 1990, Donald 1991, Hutchins 1995, Knorr Cetina 1999, Giere 2002, Strevens 2003, Muldoon and Weisberg 2011). The simplest way to show this is in aggregative terms. Suppose that Stan and Oliver aim to accomplish a feat of memory – let's say remembering all the numbers in a telephone book. One way they might do this would be for each to start at 'a' and try to get through as many as possible. This way, they will both remember the same entries, assuming they have similar cognitive capacities. Another, however, would be for Stan to start at 'a', and Oliver to start at 'z', working backwards. This way they will individually remember just as many entries, but not the same ones (assuming neither gets past half way in the book). But as long as they are willing to share information with one another, the second method allows them *as a pair* to store twice as much knowledge as the first does.

Obviously something like this kind of distribution of labor is at work in linguistic communities. Some individuals specialize in medicine, and devote their efforts to storing up medical knowledge. Others devote themselves to botany. When the doctor needs to know something about plants, she asks a botanist; and when the botanist needs to know something about medicine, he asks the doctor. The result is that the total knowledge capacity of the community increases, and arguably this marks a major evolutionary step forward in our species' history (Donald 1991: 308-312; cf. Tomasello 1999: 37ff). Characterized in these terms such a distribution of cognitive labor would be a distribution of propositional knowledge – the doctor stores knowledge about arthritis, the botanist stores knowledge about trees. The aggregation of the individual pieces of knowledge creates a greater overall store, but each piece is fully functional on its own. Semantic deference, however, holds that we also have a distribution of semantic knowledge. But we might suspect that dividing the labor of storing semantic knowledge can optimize a community's capacity to store knowledge even further.

Suppose a doctor and a horticulturalist, Delilah and Harriet, recognize that there is a great deal of recorded knowledge concerning herbal remedies – they have found, let's say, *Bartram's Encyclopaedia of Herbal Medicine*. They decide between them to store as much of this knowledge as they can in their memory. But, Delilah recognizes, while her grasp of medical concepts is in great shape – she can identify thousands of medical conditions – she has a very poor grasp of plant-concepts, barely being able to tell an Elm from an Oak. Harriet, conversely, has thousands of horticultural concepts stored in her biological memory, but few if any medical concepts – she thinks arthritis is a disease of the muscles, and narcolepsy a disease of the nose. Every entry in *Bartram's*, however, involves both a medical and a horticultural concept. What is the most efficient way for the two to remember the contents of the Encyclopaedia? To begin with, they can each start at opposite ends and agree to share information, as Stan and Oliver did with the phone book. But clearly it will optimise the process further still if they don't bother to try to acquire the semantic knowledge that underpins terms in the entries that they don't already have. Instead, the doctor simply learns off the entries starting at 'a' without grasping the horticultural concepts in those entries, and the horticulturalist starts at the other end of the book, learning the entries by rote but without grasping the meaning of the medical terms. Stan and Oliver doubled the amount of entries they were able to store by starting at opposite ends of the phone book and cooperating later. But, assuming the storage of semantic knowledge for any concept takes roughly the same amount of cognitive space, and that an entry with two concepts takes about twice the cognitive capacity to remember as an entry with one, Delilah and Harriet have now *quadrupled* the number of entries they can memorize as a pair. If all the copies of *Bartram's* are someday burned, the strategy adopted by this pair offer us our greatest hope of retaining the knowledge within.

Clearly we do not store knowledge in our linguistic communities by jointly learning off the contents of Encyclopaedias. But the case illustrates the point: by dividing the labour of storing semantic knowledge, and thereby eliminating redundancy in the storage of that knowledge, we can dramatically increase our ability to store knowledge as a community. I think all the signs are there that our communities do employ this strategy. For example, doctors are skilled at identifying medical problems, and they often prescribe courses of medicine to patients that they have learned are appropriate for those conditions from a book. But they are often not in a position to identify the medicines themselves – this is the job of the pharmacist. There is no need for the doctors to store the semantic knowledge behind the various pharmaceutical terms that they use to pick out the medicines they prescribe, because they know the pharmacist is in possession of the semantic knowledge for those terms. Importantly, with an account of how beliefs that involve terms that are not fully understood by the bearer of the belief can be fully functional, we can now see how the doctor has fully functional beliefs that involve pharmaceutical concepts, without needing to go to the trouble of locally storing the semantic knowledge behind those concepts.

As a result, we how have an answer to the question *why* we might employ deference: although particular deferential beliefs require collaborators to be fully implemented and hence are less efficient than particular non-deferential beliefs, the total amount of concepts and thereby knowledge that a community is

in a position to store if it employs deference is far greater than it will be without that strategy. While the utility of specific deferential beliefs is lower than non-deferential beliefs, then, the utility of a network of deferential beliefs is far higher, since as a community we are in a position to store far more knowledge with such a network than without.

But there is one final point to add. Recall that the first and third of the 'Glue and Trust' conditions, that the knowledge is typically invoked and automatically endorsed, automatically apply in cases of deference. The fourth doesn't really make sense in the case of semantic knowledge. And the second, while indeed putting a brake on the range of propositional knowledge that might be stored 'offline', is not in fact a constraint on the semantic knowledge that can underpin a well functioning belief, which can be greatly delayed without undermining the functionality of the belief. All this means that the extent to which semantic knowledge can be 'outsourced' is *far less constrained* than the extent to which propositional knowledge can be. Although Stan and Oliver may have managed together to store twice as much information in the phone book as they might have done without cooperating, it is not at all obvious that Stan *knows* the entries that Oliver has remembered, or vice versa. They know where to look them up, but as discussed, simply knowing how to retrieve the information is not sufficient to show that we already know it. As a pair, then, they have retained twice as much information using this strategy, but individually, they have no more knowledge than they would have had if they had each learned all the same entries. Delilah and Harriet, on the other hand, each know the entries that they have memorized, even though the semantic knowledge for at least once concept in each proposition they have learned is stored in their partner's memory. Each propositional memory they have formed satisfies *Utility*, and if we suppose the information is reliable and acquired in such a way as to count as knowledge, as many suppose beliefs acquired through testimony can (Lackey 2007, Pritchard 2010), then it would seem that they know the entries they have stored. But as we saw, because they are outsourcing the semantic knowledge behind half the terms in their beliefs, while they are quadrupling the total amount of knowledge retained as a pair, they are already individually doubling their store of knowledge – something Stan and Oliver were unable to do. The Glue and Trust conditions, in other words, by identifying sound constraints on the functional efficacy of a piece of knowledge, greatly restrict the extent to which we can plausibly outsource propositional knowledge; but they do not constrain the outsourcing of semantic knowledge to anything like the same degree.

The outsourcing of semantic knowledge, as a result, is potentially the site of a far greater expansion of our epistemic capacity than the outsourcing of propositional knowledge. I can acquire knowledge of propositions like 'penicillin kills bacteria', 'quarks have the same spin as antiquarks', 'there's a black hole relatively nearby', or 'genes are replicators', all without having the ability to identify the reference of at least one of the terms in each of these, by outsourcing the semantic knowledge behind them. Some have argue that extended mind reasoning about the outsourcing of propositional knowledge leads to an 'explosion of knowledge', including the entire contents of the internet (Ludwig 2015), but this comes at the price of undermining the Glue and Trust conditions, and with it the plausibility of the claim. The outsourcing or semantic knowledge, however, seems to already permit a dramatic expansion of our epistemic

capacity without undermining those conditions. Not only does it turn out that semantic deference is best made sense of in terms of extended mind reasoning, then, but some of the more tantalizing consequences of extended mind reasoning are better made sense of when that story is told in terms of semantic deference.

5. Conclusions

Semantic deference is sometimes thought of as a sort of an epiphenomenon of language-based thought – a third wheel that, although certain compelling examples from Burge and others leave us with no choice but to accept, seems to force upon us an unwieldy metalinguistic account of the content of our beliefs and do no real work in our cognitive lives. When viewed through the lens of extended mind reasoning, however, deferential thoughts can be accounted for in a way that avoids the problems of a metalinguistic account, and can be shown to optimize rather than undermine the efficiency of our cognitive lives. When we divide the labour of storing concepts, the number of concepts we are in a position to use increases dramatically, and with it the overall knowledge bearing capacity both of the community and the individual.

Allen-Hermanson, S. (2012). Superdupersizing the mind: Extended cognition and the persistence of cognitive bloat. Philosophical Studies, 158(1), 1–16.
Bartram, T. (1998). Bartam's Encyclopedia of Herbal Medicine. Little, Brown Book Group.
Brown, J. (1998). Natural Kind Terms and Recognitional capacities. Mind, 107: 275– 304.
Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior, 5, 325 – 337.
Burge, T. (1979). Individualism and the Mental. In French, Uehling, and Wettstein (eds.) Midwest Studies in Philosophy IV, University of Minnesota Press, pp. 73-121.
Chalmers, D. (2002a). On Sense and Intension. Philosophical Perspectives, 16: 135-82.
Chalmers, D. (2002b). The Components of Content. In David Chalmers (ed.) The Philosophy of Mind: Classical and Contemporary Readings. Oxford: Oxford University Press.
Chalmers, D. (2006). The Foundations of Two-Dimensional Semantics. in Two-Dimensional Semantics: Foundations and Applications, ed. M. Garcia-Caprintero and J. Macia, Oxford University Press.
Clark, A. (2003). Natural-born cyborgs. New York: Oxford University Press.
Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension . Oxford: Oxford
Clark, A., & Chalmers, D. (1998). The Extended Mind. Analysis, 58 , 7-19.
Crane, T. (1991). All the Difference in the World. Reprinted in S. Goldberg and A. Pessin (1996) The Twin Earth Chronicles. New York: M. E. Sharpe. 284-304.
Dokic, J. (2015). Seeds of Self-Knowledge: Noetic Feelings and Metacognition. In Foundations of Metacognition, Eds Michael Beran, Johannes Brandl, Josef Perner and Joelle Proust, Oxford University Press.
Donald, M (1991). Origins of the Modern Mind.
Evans, G. (1983). Varieties of Reference. Oxford University Press.

Fodor, J. (1998a). Concepts. Oxford: Oxford University Press.
Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34, 906 - 911.
Giere, R. (2002). Distributed Cognition in Epistemic Cultures. Philosophy of Science 69: 637–44.
Goldberg, S. (2009). Experts, Semantic and Epistemic. Nous.
Gopnik, A. & Meltzoff, A.N. (1997). Words, Thoughts, and Theories. Cambridge, MA: Harvard University Press.
Hutchins, E. (1995). Cognition in the Wild. MIT Press.
Jackson, F. (1998). From Metaphysics to Ethics. Oxford: Oxford University Press.
Jackson, F. (1998b). Reference and Description Revisited.
Kitcher, P (1990). The Division of Cognitive Labor. Philosophy of Science 87: 5–22.
Knorr Cetina, K. (1999). Epistemic Cultures: How the Sciences Make Knowledge. Cambridge, MA: Harvard University Press.
Koriat, A (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. Journal of Experimental Pyschology: General, 123, 311-33.
Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. Consciousness and Cognition, 9, 149-171.
Kripke, S. (1980). Naming and Necessity. Harvard University Press.
Lackey, J. (2007). Why We Don't Deserve Credit for Everything We Know. Synthese 158, 345-61.
Kriegel, U. (2013). The Phenomenal Intentionality Research Program. In Uriah Kriegel and Terrence Horgan (eds.) Phenomenal Intentionality. Oxford University Press.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. Psychological Review, 85: 207 - 238.
Mendelovici, A. (2010). Mental Representation and Closely Conflated Concepts. Dissertation, Princeton University.
Muldoon, R. and Weisberg. M. (2011). Robustness and Idealization in Models of Cognitive Labor. Synthese 183.2: 161–74.
Peacocke, C. (1992). A Study of Concepts. MIT Press.
Pritchard, D. (2010). Cognitive Ability and the Extended Cognition Thesis. Synthese August 2010, Volume 175: 133-151.
Prinz, J. and Clark, A. (2004). Putting Concepts to Work: Some Thoughts for the Twenty-first Century. Mind & Language, Vol. 19 No. 1 February: 57–69.
Putnam, H. (1975). The meaning of 'meaning'. In Putnam, H. (1975) Philosophical Papers, ii. Cambridge: Cambridge University Press.
Putnam, H. (1963). Brains and Behavior. Analytical Philosophy, Second Series, ed. R. J. Butler (Oxford: Basil Blackwell): 211-235.
Rosch, E. and Mervis, C.B. (1975). Family Resemblance: Studies in the Internal Structure of Categories. Cognitive Psychology 7: 573-605.
Rupert, R. (2009). Cognitive systems and the extended mind. Oxford: Oxford University Press.
Russell, B. (1921). The Analysis of Mind. Allen and Unwin.
Schwartz, B. and Metcalf, S. (2010). Tip-of-the-tongue (TOT) states: retrieval, behavior, and experience. Journal of Memory and Cognition,
Searle, J. R., (1991). Consciousness, Unconsciousness, and Intentionality,"

Philosophical Issues, 1, 45-66.
Searle, J. R. (1992). The Rediscovery of Mind, Cambridge, MA: M.I.T. Press.
Soames, S. (2002). Beyond Rigidity: The Unfinished Program of Naming and Necessity. Oxford: Oxford University Press.
Soames, S. (2005). 'Reference and Description.' In F. Jackson and M. Smith (eds.) The Oxford Handbook of Contemporary Philosophy. Oxford: Oxford University Press. 397- 426.
Sterelny, K. (1983). Natural Kind Terms. Pacific Philosophical Quarterly, 64: 110-125.
Strevens, M. (2003). The Role of the Priority Rule in Science. Journal of Philosophy 100.2 (2003): 55–79.
Tollefsen, D. (2006). From Extended Mind to Collective Mind. Cognitive Systems Research, 7: 140-150.
Tomasello, M. (1999). The Cultural Origins of Human Cognition. Harvard University Press.
Varga, S. (2013). Frames of cognition. Cognitive Systems Research, 25-26, 54–60.
Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), Theories of group behavior (pp. 1 85-208). New York: Springer-Verlag.
Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. Psychological Review , 86(1):44-60.
Wickelgren, W. A. (1992). Webs, cell assemblies, and chunking in neural nets. Concepts in Neuroscience , 3(1):1-53.